

Thomas Klein

# Mittelhochdeutsche Wortfamilien: Ermittlung und Perspektiven

**Abstract:** An inventory of the Middle High German word families is still missing whereas the Old High German and New High German word families are recorded by the dictionaries of J. Splett. In this paper a semi-automatic method is represented which can help to find and analyze the Middle High German word families. By several scripts a combined list of MHG and OHG lemmata is transformed and expanded to a table containing among other things a column with a simplified variation of Splett's word formation formulas and a column with the common base of the word family the lemma probably belongs to. In a labour-intensive last step, these proposals have to be manually checked and corrected.

## 1. Mittelhochdeutsche Wortbildung: Stand und Aufgaben

Zwar gibt es eine neuere Gesamtdarstellung der mhd. Wortbildung (Klein, Solms, und Wegera 2009) und eine Reihe wichtiger und gewichtiger Monographien zur Wortbildung der einzelnen Wortarten (Herbers 2002; Leipold 2006; Ring 2008; Ganslmayer 2012). Doch bleibt hier durchaus noch einiges zu tun. Die gravierendste Lücke scheint mir zu sein, dass die mhd. Wortfamilien noch weithin im Dunkeln liegen. Weder ihre ungefähre Zahl noch Umfang und Struktur der einzelnen Wortfamilien sind bekannt, noch lassen sich bisher Aussagen machen über Wachstum und Schrumpfen von Wortfamilien innerhalb der mhd. Zeit oder im Vergleich mit dem Althochdeutschen und Neuhochdeutschen.<sup>1</sup> Für diese Sprachepochen sind die Wortfamilien von Jochen Splett bereits umfassend ermittelt und dokumentiert worden (Splett 1993; 2009).

Spletts Wörterbücher bieten

- eine strukturierte Darstellung aller Wortfamilien; die Binnenstruktur der Wortfamilie wird durch eine dezimale Gliederung verdeutlicht (s. Abb. 1);
- die Analyse der Wortbildungsstruktur aller Wörter in Gestalt von „Strukturformeln“, wobei von der Form, Funktion und Semantik der beteiligten Affixe und Kernwörter abgesehen wird.

---

1 Angeregt durch zwei Vorträge, die ich dazu und zur Verknüpfung digitaler Lemmalisten historischer Sprachstufen des Deutschen 2013 und 2016 bei den „Arbeitsgesprächen zur historischen Lexikographie“ in Bullay (<https://www.uni-trier.de/index.php?id=14673>) gehalten habe, gibt es erfreulicherweise im Projekt ZHistLex („eHumanities-Zentrum für Historische Lexikographie“) ein Arbeitspaket, das den „Epochenübergreifenden Wortfamilienstrukturen“ in ahd. und mhd. Zeit gewidmet ist, vgl. <http://zhistlex.de/ziele/arbeitspakete/>.

Abb. 1: Wortfamilie *ahd. BREMAN* in Splett (1993, I.1: 101).

BREMAN (1)		
1.0	<b>breman</b> ** st.V.	wV/ 'brüllen'
1.1.3	<b>brāmen</b> * sw.V.	(wV)Vjan/ 'brüllen'
1.1.4	<b>brummen</b> * sw.V.	(wV)Vjan/ 'brünstig grunzen'
1.2.1	<b>brema</b> * sw.M.	(wV)San/ 'Bremse [Insekt]'
	<b>brema-krūt</b> ? ** st.N.	((wV)San) (wS) / 'Taubnessel' (2)
	<b>-stal</b> * st.M.	/ 'Sommergehege [Ort zum Schutz der Herde gegen Bremsen]'
	<b>bremen-krūt</b> ** st.N.	/ 'Taubnessel' (2)
	<b>brema</b> * sw.F.	(wV)Són/ 'Bremse [Insekt]'
1.2.2	<b>bremma</b> * F.	(wV)S / 'Bremse [Insekt]' (3)
2.2.1	<b>brim-issa</b> * F.	((wV)San) sS / 'Bremse [Insekt]' (4)
2.3.4	<b>brum-ft-ig</b> * Adj. <sup>2</sup>	[(wV)sS]sA [ ((wV)Vjan) sS]sA/ 'brünstig'

Zu einer „Grundlegung einer zukünftigen Strukturgeschichte des deutschen Wortschatzes“, die der Untertitel von Spletts Wörterbüchern (Splett 1993; 2009) als Ziel benennt, wäre aber noch eine vergleichbare Erfassung der mhd. und frühhd. Wortfamilien wünschenswert. Der dafür nötige Arbeitsaufwand ist jedoch immens. Es ist aber möglich, ihn ausgehend von Spletts Wortfamilienwörterbüchern und mit (semi)automatischen Verfahren ein gutes Stück zu verringern. Dies soll im Folgenden für das Mittelhochdeutsche gezeigt werden. Schon vorab aber ist zu betonen, dass die hier vorgeschlagene Verfahrensweise eine linguistische „händische“ Wortfamilienklassifikation nicht ersetzen, sondern nur erheblich erleichtern kann.

## 2. Voraussetzungen und Vorarbeiten für die Ermittlung mittelhochdeutscher Wortfamilien

### 2.1. Althochdeutsche und mittelhochdeutsche Lemmalisten

Es liegt nahe, für eine Ermittlung der mhd. Wortfamilien Spletts Wortfamilienwörterbücher als Muster und Ausgangspunkt zu nutzen. Voraussetzung dafür ist allerdings, dass diese Wörterbücher in digitaler Form verfügbar sind. Für den Wortindex zu Spletts Althochdeutschem Wörterbuch (Splett 1993) ist das der Fall. Auf der mhd. Seite gibt es die Bonner Lemmaliste des Projekts „Mhd. Grammatik“. Im Umfang deckt sich diese Liste weitgehend mit dem Lexembestand von Lexers Mhd. Wörterbuch, vermehrt um den Bestand des Trierer Findebuchs (Gärtner et al. 1992). Die Lexemansätze der Bonner Lemmaliste entsprechen denen des Korpus der Mhd. Grammatik (MiGraKo). Diese Ansätze orientieren sich jedoch stärker als die Lexers oder gar des MWB an grammati(kographi)schen Bedürfnissen und sind daher auch für Untersuchungen zur Wortbildungsstrukturen weit geeigneter. Mit Lexers Wörterbuch und

dem Findebuch teilt die Bonner Lemmaliste andererseits den Nachteil, viele Lexeme zu enthalten, die ausschließlich in frühneuhochdeutschen Quellen oder in Quellen aus der spätmittelhochdeutsch-frühmittelhochdeutschen Übergangszeit belegt und damit für den engeren mhd. Zeitraum vom späten 11. bis zur Mitte des 14. Jahrhunderts nicht gesichert sind. Besser wäre es daher, diese Lexeme von einer Untersuchung der mhd. Wortfamilien auszuschließen. Für die bisher erschienene Alphabetstrecke (*a – hanfken*) sollte dafür der Lexembestand des MWB maßgeblich sein. Für die Erprobung des hier vorgestellten Verfahrens musste jedoch noch die unveränderte Bonner Lemmaliste zugrunde gelegt werden.

Von den Lexemansätzen der mhd. Wörterbücher weichen die internen Ansätze des Projekts „Mhd. Grammatik“ und damit auch der Bonner Lemmaliste aus grammatikographischen Gründen in folgenden Punkten ab:

- Die Ansätze sind frühmhd., also ohne Auslautverhärtung, ohne Lenisierung *-nt- > -nd*, z. B. *bald, lib, tag, winten, winden*‘.
- Anders als auch in den ahd. Wörterbüchern werden Geminaten im Auslaut nicht gekürzt, z. B. *ball, sinn, spott, schiff*.
- Die mhd. Schwa-Apokopierungen und Synkopierungen werden nur insoweit berücksichtigt, als das „normalmhd.“ getilgte Schwa in Rundklammern erscheint, z. B. *ar(e)* ‚Adler‘.
- Bei den kurzen e-Lauten werden unterschieden:
  - *ë* für altes (germ.) *e*,
  - *è* für (Primär- und Sekundär-)Umlaut-*e*, als Diphthongkomponente in *iè* und *üè* und in *oè* für langes *œ*,
  - *e* steht ausschließlich für Schwa.
- Die Lexeme sind wie bei Lexer morphologisch segmentiert, anders als bei Lexer aber konsequent auch in Fällen wie *ver-ant-würten* (Lexer *ant-würten*, aber *ver-antwürten*). Abgetrennt werden wie bei Lexer nur die „neuen“ Suffixe wie *-heit, -schaft, -tuom*, nicht auch die „alten“ wie *-ig, -(e)de < (-ida, -idi), -inne* usw.

## 2.2. Verknüpfte Lemmaliste Ahd. – Mhd.

Für interne Zwecke des Projekts Mhd. Grammatik wurde bereits vor mehr als 10 Jahren in Bonn der „Alphabetische Wortindex“ aus Spletts Ahd. Wörterbuch (Splett 1993 II: 389–620) mit der Bonner Lemmaliste verknüpft. Dazu setzte ein Script die ahd. Lexeme provisorisch ins Mhd. um und fügte das künstlich erzeugte mhd. Lexem jeweils dem Datensatz des ahd. Lexems hinzu. Sodann prüfte das Script, ob das provisorische mhd. Lexem in der mhd. Bonner Lemmaliste enthalten ist; wenn nein, wurde es mit einem initialen Stern markiert. Jedoch konnten so nur bei weniger als 40% der ahd. Lexeme die mhd. Nachfolger automatisch zugeordnet werden. Dies liegt in

der Hauptsache daran, dass sich nur knapp die Hälfte des ahd. Wortschatzes noch im Mhd. belegen lässt. Außerdem wich in nicht wenigen Fällen die automatisch erzeugte mhd. Form von der Gestalt des mhd. Nachfolgelexems in der mhd. Bonner Lemmaliste ab, sodass die Zuordnung unterblieb. Aus diesem Grund musste insbesondere bei allen Fällen mit sternmarkiertem mhd. Lexem geprüft werden, ob nicht doch ein mhd. Nachfolgelemma existiert. Das Ergebnis dieser Arbeiten ist eine verknüpfte ahd.-mhd. Lemmaliste (s. Tab. 1).

Tab. 1: Ausschnitt aus der Bonner Lemmaliste Ahd. – Mhd.

ahd. Lexem	Beleglage	mhd. Lexem <sup>1)</sup>	ahd. Wortfamilien- kernwort
bald, Adj.	(+*)	bald, adj	BALD
balda, st.F.[(j)ô]	*	balde, f	BALD
baldên, sw.V.	(*)	balden, swv	BALD
bald-lîh, Adj.	*	bald-lich, adj	BALD
bald-lîhho, Adv.	(*)	bald-lîche, adv	BALD
bald-lîhhûn, Adv.	*	bald-lîche, adv	BALD
baldo, Adv.	(+*)	balde, adv	BALD
belden, sw.V.	(+*)	bêlden, swv	BALD
beldî, st.F.[î]	(*)	bêlde, f	BALD
beldida, st.F.[ô]	(*)	*beldede, f	BALD
beldirîh, st.M.[a/i]	*	balderich, m	BALD
fra-bald, Adj.	*	*vra-bald, adj	BALD
fra-bald-lîh, Adj.	*	*vra-bald-lich, adj	BALD
fra-bald-lîhho, Adv.	*	*vra-bald-lîche, adv	BALD
fra-baldo, Adv.	*	*vra-balde, adv	BALD
fra-beldî, st.F.[î]	*	*vra-bêlde, f	BALD
gi-bald, Adj.	*	*ge-bald, adj	BALD
gi-baldôn, sw.V.	*	ge-balden, swv	BALD
ir-baldên, sw.V.	(*)	er-balden, swv	BALD
ir-baldnissi, st.N./F.[jaî]	*	er-baldnisse, f	BALD
ir-baldôn, sw.V.	()	er-balden, swv	BALD
ir-belden, sw.V.	*	*er-bêlden, swv	BALD
un-bald, Adj.	()	*un-bald, adj	BALD
un-baldo, Adv.	+	*un-balde, adv	BALD
un-beldî, st.F.[î]	()	*un-bêlde, f	BALD
bora-baldo, Adv.	()	*bore-balde, adv	BALD + BOR
kurzi-bold, st.M.[a/i]	*	kurze-bold, m	BALD + KURZ

<sup>1)</sup> \* vor dem Lemma = mhd. nicht belegt, automatisch konstruiert

In die mhd. Bonner Lemmaliste wurde dann dort, wo für ein mhd. Lexem sein ahd. Vorläufer belegt ist, dieses ahd. Lexem nebst grammatischen Angaben und das zugehörige Wortfamilienkernwort aus Spletts Wortindex inseriert (s. Tab. 2).

Statistisch gesehen deutet die ahd.-mhd. Lemmaliste auf ein beträchtliches Maß an lexikalischer Diskontinuität: Nur für knapp die Hälfte der ahd. Lexeme (ca. 47%) finden sich Nachfolger in der Bonner Lemmaliste und nur etwa ein Sechstel (ca. 17%) der mhd. Lexeme dieser Liste haben Vorläufer, die in ahd. Quellen belegt sind. Zu bedenken ist allerdings, dass in der vergleichsweise spärlichen ahd. Überlieferung sehr viele ahd. Lexeme, die sich im Mhd. fortsetzen, zufällig nicht belegt sein dürften. Zudem sind in die Bonner Lemmaliste viele Lexeme aus Lexer und dem Findebuch (Gärtner et al. 1992) übernommen, die erst in frühmhd. Zeit belegt sind und daher eigentlich nicht einbezogen werden dürften (s. oben 2.1). Andererseits enthält die Liste auch Wortgut, das ausschließlich in Quellen der ahd.-frühmhd. Übergangszeit belegt ist. Berücksichtigt man diese Faktoren, so dürfte der tatsächliche Anteil der mhd. Lexeme mit ahd. Vorgeschichte erheblich höher liegen als 17%.

Nicht ganz selten sind zwei ahd. Lexeme ausdrucksseitig in einem mhd. Lexem zusammengefallen. Die Regel ist das bei den ahd. *-ên-* und *-ôn-* Verben, die im Ahd. oft auch semantisch weitgehend übereinstimmen; vgl. z. B. ahd. *ir-baldên, ir-baldôn* ‚sich erkühnen, wagen‘ → mhd. *er-balden* (s. Tab. 1). In der Bonner Lemmaliste sind dann dem mhd. Lexem beide ahd. Vorgängerlexeme zugeordnet (s. Tab. 2). Ähnlich wurde verfahren, wenn zwei semantisch verschiedene ahd. Lexeme im Mhd. ausdrucksseitig zusammengefallen sind und so zu einem polysemen mhd. Nachfolgelexem geführt haben, vgl. *bleichen* in Tab. 2.

Tab. 2: Ausschnitte aus der erweiterten Bonner mhd. Lemmaliste.

<i>MiGraKo</i> Lexem POS	<i>(Lexer/MWB)</i> MWB-ID	<i>Althochdeutsch</i>	
		Lexem POS	Wortfamilien- Kernwörter
bal-rât m<u>	bal-rât stm 10836000	balo-rât st.M.	BALO + RâTAN
bal-rêdære m	bal-reder stm 10842000	–	
...	...	...	...
balas m	balas stm 10683000	–	
bald adj	balt a 10959000	bald Adj.	BALD
bald-hêt f	balt-heit stf 10965000	–	
bald-lich adj	balt-lich a 10968000	bald-lih Adj.	BALD

<i>MiGraKo</i>	<i>(Lexem/MWB)</i>	<i>MWB-ID</i>	<i>Althochdeutsch</i>	
Lexem POS			Lexem POS	Wortfamilien-Kernwörter
bald-lîche adv	balt-lîche adv	10974000	bald-lîhho Adv./ bald-lîhhûn Adv.	BALD
bald-lôs adj	balt-lôs a	10977000	–	
bald-spræche adj	balt-spræche a	10980000	–	
bald-spræche f	–	–	–	
balde adv	balde adv	10704000	baldo Adv.	BALD
balde f	balde stf	10707000	balda st.F.	BALD
baldekîn mn	baldekîn stm	10719000	–	
...	...	...	...	...
blëichen swv {1} ‚bleich werden‘, {2} ‚bleich machen‘	bleichen swv	22158000	{1} bleihhên sw.V., {2} bleihhen sw.V.	BLÏHHAN
...	...	...	...	...
er-balden swv	er-balden swv	40434000	ir-baldên sw.V., ir-baldôn sw.V.	BALD

### 2.3. Übertragung der Strukturformeln auf die mhd. Nachfolgelexeme

Wenn Spletts Ahd. Wörterbuch komplett digitalisiert vorliegt, so lassen sich auch die Strukturformeln von den ahd. Lexemen auf ihre mhd. Nachfolger übertragen. Dazu müsste ein Script lediglich die ahd. Lexeme aus der vierten Spalte der Bonner Lemmaliste im Digitalisat von Splett Bd. I,1–2 aufsuchen, dort die Strukturformel entnehmen und in einer zusätzlichen Spalte der Bonner Lemmaliste einfügen. Ebenso könnten die Angaben zur Binnenstruktur der Wortfamilie in Gestalt der Dezimalgliederung (s. Abb. 1) übernommen werden. In den Strukturformeln wären die Stammbildungsangaben, die für das Mhd. nicht mehr von Belang sind, zu löschen, z. B. mhd. *balden* swv → ahd. *baldên* → ahd. Strukturformel (wA)Vên → mhd. (wA)V. Wo ein mhd. Lexem mehrere ahd. Vorgänger hat, werden deren Strukturformeln dadurch in der Regel zusammenfallen, z. B. mhd. *er-balden* swv → ahd. *ir-baldên*, *ir-baldôn* → ahd. Strukturformel p(wA)Vên, p(wA)Vôn → mhd. p(wA)V, p(wA)V → p(wA)V.

Für die Frage, welche Teile der ahd. Wortfamilien sich im Mhd. fortsetzen, würde sich dagegen die umgekehrte Blickrichtung empfehlen: Ausgehend von Spletts Wörterbuch würden in der ahd.-mhd. Lemmaliste die mhd. Nachfolgelexeme gesucht und erweiterten Datensätzen des Wörterbuchs hinzugefügt. Ist

das mhd. Lexem *besternt* (also nicht belegt, sondern nur konstruiert), so wäre auch das ahd. Lexem entsprechend zu markieren (etwa durch ein Sterbekreuz). Man würde so auch sehen, in welcher Weise die Binnenstruktur der Wortfamilien (in Gestalt der dezimalen Gliederung in Spletts Wörterbuch, s. Abb. 1) durch die Lexemverluste beeinträchtigt ist. Da im Schnitt nur die Hälfte der belegten ahd. Lexeme im Mhd. überlebt hat, dürften diese „Erosionsschäden“ teils beträchtlich sein.

Für die mhd. Seite wäre durch diese Übertragung der Wortbildungsstrukturdaten aus dem Splett-Wörterbuch zunächst wenig gewonnen. Denn erfasst werden damit nur die 17% der mhd. Lexeme, die einen belegten ahd. Vorläufer haben. Es fragt sich also, wie auch für den gesamten Rest mit vertretbarem Aufwand eine Zuordnung zu Wortfamilien erreichbar wäre. Im Folgenden wird ein semiautomatisches Verfahren vorgestellt, das diese Zuordnung erleichtert und dabei zugleich (vereinfachte) Strukturformeln konstruiert.

### 3. Konstruktion von mhd. Strukturformeln und Kernwortextraktion

#### 3.1. Übersicht

Der automatische Teil des Verfahrens soll zu jedem mhd. Lexem eine (vereinfachte) Strukturformel erzeugen und die in ihm enthaltenen ‚Kernwörter‘<sup>2</sup> (Grundmorpheme) ihrer Wortfamilie (kurz: WF) zuweisen. In der Übersicht am Beispiel *bēte-gültig* ‚steuerpflichtig‘:

Ausgangslexem	bēte	– gültig	
	↓	↓	
Segmentierung (Suffixabtrennung)	bēt-e	– gült-ig	
	↓	↓	
Extraktion u. Strukturformelerzeugung	bēt	gült	→ wXlsS+wXlsA
	↓	↓	
Umlautbereinigung	bēt	gult	
	↓	↓	
Alternanzbereinigung	bit	gēlt	
	↓	↓	
Sortieren	WF	WF	
	BITTEN	GELTEN	

2 Dieser Terminus nach Splett (1993: XXIII).

## 3.2. Vorbereitende Schritte

### 3.2.1. Suffixabtrennung

In der Bonner Lemmaliste sind im Gefolge Lexers nur die Präfixe und die noch erkennbar aus Kompositionsgliedern entstandenen „Neu-Suffixe“ wie *heit*, *schaft*, *bære*, *haft* usw. durch „-“ abgetrennt und außerdem Kompositionsfugen ebenso bezeichnet. Für eine Extraktion der selbst möglichst nicht mehr suffigierten Kernwörter ist aber auch die Abtrennung der „Alt-Suffixe“ wie z. B. *ig*, *ing*, *inne*, *nisse*, *e*, *en*, *ede* usw. erforderlich, z. B. *aller-diè-müètigest* → *aller-diè-müèt-ig-est* (*müèt* ist hier umgelautete Variante des Kernworts *muot*).

Diese Abtrennung der „Alt-Suffixe“ ist einer der kritischsten Punkte des gesamten Verfahrens, denn es ist unvermeidlich, dass es dadurch auch zu unerwünschten Kernwort-Homonymien kommt. Dieser Schaden sollte durch a) eine optimale Auswahl der abzutrennenden Suffixe und b) eine geeignete Formulierung der Abtrennungsregeln aber so klein wie möglich gehalten werden. Beispiele:

- Wenn etwa *er* global abgetrennt würde, so wäre der Schaden vermutlich größer als der Gewinn, da die Zahl unerwünschter Kernwort-Homonymien sehr groß würde, vgl. z. B. (\*⇒ steht für falsche Wortfamilienzuordnung):

*muoter* → *muot-er* \*⇒ Wortfamilie MUOT; *meister* → *meist-er* \*⇒ MEIST; *bitter* → *bitt-er* \*⇒ BITTEN; *laster* → *last-er* \*⇒ LADEN usw.

Daher wird bei der aktuellen Script-Version auf eine Abtrennung von *er* ganz verzichtet.

- Von einer globalen Suffixabtrennung können bestimmte Kernwörter oder Kernwortvarianten vorab durch ihre Großschreibung in der Lemmaliste ausgenommen werden, da das Script Groß- und Kleingeschriebenes strikt unterscheidet, z. B. (-)*adel*(-) → (-)ADEL(-), (-)*èdel*(-) → (-)ÈDEL(-); ADEL ist dann von der automatischen Abtrennung des Suffixes *-el* nicht mehr betroffen.
- Es darf nicht abgetrennt werden, wenn das danach links verbleibende Segment nicht silbisch ist, z. B. *bringen* → *bring-en*, nicht → **br-ing-en**.
- Wird ein mit *i* anlautendes Suffix abgetrennt, so darf das links verbleibende Segment nicht auf *e* enden, da sonst der Diphthong *ei* zertrennt würde, z. B. *abe-neig-en* nicht → *abe-ne-ig-en*.
- Wichtig auch: Suffixe dürfen nicht ihrerseits noch einmal segmentiert werden, weil sie dann später nicht mehr als Suffix erkannt werden, also z. B. *burg-mæze* ‚burgmässig‘ nicht → \* *burg-mæz-e*. Das hat bei Homonymie von Suffix und Kernwort(variante) (s.u. 3.2.3) den Nachteil, dass der Segmentierungsschritt auch bei den (wenigen) Bildungen mit dem Kernwort

unterbleibt und damit auch die Kernwortextraktion und Zuordnung zur Wortfamilie nicht zum Ziel führt, z. B. nicht *un-mæze* → *un-mæz-e* → *mæz* → *mâz* → *mëz* ⇒ MËZZEN.

### 3.2.2. Affixliste

Wenn festgelegt ist, was als Affix zu betrachten ist, werden Präfixe und Suffixe in einer zweispaltigen Liste zusammengefasst, in der den Affixen jeweils das Symbol zugeordnet ist, durch das sie in der Strukturformel vertreten sind. Für Präfixe ist das „p“, für Suffixe „s“. Wenn das Suffix wortarteindeutig ist, wird das Wortartsymbol hinzugefügt: „sS“ Substantiv-, „sA“ Adjektiv-, „sAD“ Adverbsuffix; wenn sich das Suffix mit mehreren Wortarten verbinden kann, sollten möglichst alle aufgeführt werden, z. B. *-în* sS, sA, *-haft* sA, sAD; alternativ kann „sX“ ohne Festlegung der Wortart zugeordnet werden. Bei genuseindeutigen Substantivsuffixen wird das Genus in eckigen Klammern hinzugefügt, z. B. *-schaft* sS[f].

Tab. 3: Beispiele aus der Affixliste.

Präfix	Struktursymbol	Suffix	Struktursymbol
be	p	ig	sA
ge	p	lich	sA
ent	p	liche	sAD
abe	p	hêt	sS[f]
ane	p	(e)de	sS
gègen	p	schaft	sS[f]
		haft	sA, sAD
		e	sX

### 3.2.3. Mehrdeutige Morpheme

Besondere Schwierigkeiten bereiten bei dem späteren Ersatz der Affixe durch ihre Struktursymbole (s. unten 3.2.5) Affixe, die mit Kernwörtern homonym sind. In den meisten Fällen wird das fragliche Morphem den ersten oder letzten Teil eines komplexen Lexems bilden. Hier gibt es algorithmisierbare Regeln für die Entscheidung zwischen Affix und Kernwort. Zunächst hilft der simple Umstand, dass Präfixe nicht den letzten Teil und Suffixe nicht den ersten Teil des Lexems bilden können. Für die Entscheidung zwischen Suffix und Kernwort lässt sich weiterhin nutzen, dass das Script der Affixliste auch die Wortart und bei Substantiven das Genus des Lexems entnommen hat:

- Stimmt die vom Suffix geforderte Wortart mit der des Lexems überein, so handelt es bei dem fraglichen Morphem um das Suffix, sonst um das Kernwort, z. B. ist *anker-haft* ‚Ankergrund‘ Substantiv, während das

Suffix *-haft* nur Adjektive oder Adverbien bilden kann; also kann *haft* in *anker-haft* nur Kernwort und nicht Suffix sein.

- Stimmt das von einem Substantivsuffix geforderte Genus mit dem des Lexems überein, so liegt das Suffix vor, sonst das Kernwort. So werden mit *-schaft* nur feminine Substantive gebildet; daher handelt es bei *schaft* in *spër-schaft* mask. ‚Speerschaft‘ um das Kernwort.
- Bliebe bei der Annahme eines Suffixes als Restwort lediglich ein Morphem übrig, das formal nur Präfix sein kann, so muss das Kernwort vorliegen, da sich sonst eine nur aus Affixen bestehende kernwortlose Bildung ergäbe. So kann es sich bei *bære* und *mæze* in *ge-bære*, *ge-mæze* nur um Kernwörter handeln, da *ge* wegen des Schwa formal nur Affix sein kann (daher nicht → \*pl<sub>s</sub>A, sondern → plwV|A). Bei Lexemen wie *un-mæze* müsste das Script dagegen wissen, dass *un* nur Präfix und nicht auch Kernwort sein kann.
- Allgemeiner kann die vorstehende Regel so gefasst werden: Unmittelbar links von einem Suffix kann kein Präfix und unmittelbar rechts von einem Präfix kann kein Suffix stehen. So kann *über* in *hōch-über-ig* nur Kernwort sein, daher nicht → \*pl<sub>s</sub>A, sondern → plwAD|sA oder wAlwAD|sA, falls *hōch* als Kernwort und nicht als gradatives Präfix (Klein, Solms und Wegera 2009: § A24) gewertet würde.

Es bleiben Fälle, die durch diese Regeln nicht abgedeckt sind. So ist bei *ant-vogel* ‚Ente(nvogel)‘ und *ant-vuoz* ‚Entenfuß‘ formal nicht entscheidbar, ob *ant* Präfix oder Erstglied eines Kompositums ist. Soweit sie bekannt sind, können Kernwörter in solchen Bildungen vorab durch Großschreibung blockiert werden (da das Script nur kleingeschriebene Affixe kennt), z. B. *ant-vuoz* → *ANT-vuoz*, das dann automatisch als wX+wX analysiert würde.

### 3.2.4. Erweiterung der Bonner Lemmaliste

Die Bonner mhd. Lemmaliste wird durch ein vorbereitendes Script um zwei Felder erweitert. Das erste Feld füllt das Script mit einer Form des Lexems (Lexem2), in der die „Altsuffixe“ abgetrennt sind. Das zweite Feld ist für das Kernwort vorgesehen und wird provisorisch mit der Ausgangsform des Lexems gefüllt, die bei Simplicia wie *hūs* oder *blint* ja bereits mit dem Kernwort identisch ist.

Tab. 4: Erweiterte Bonner Lemmaliste.

	Mittelhochdeutsch		Lexer	Althochdeutsch
Lexem, POS	Lexem2	Kernwort	...	Lexem, POS Kernwort
<i>ab-grünt-lich</i> , adj	<i>ab-grünt-lich</i>	<i>ab-grünt-lich</i>	...	–
<i>runs(e)-âder(e)</i> , f	<i>runs-(e)-âder-(e)</i>	<i>runs(e)-âder(e)</i>	...	
<i>bête-gültig</i> , adj	<i>bët-e-gült-ig</i>	<i>bête-gültig</i>	...	

### 3.2.5. Konstruktion der mhd. Strukturformel und Kernwortextraktion

Aus den Segmenten von Lexem2 erzeugt das Hauptsript mittels der aus der Affixliste bezogenen Informationen Strukturformeln, die gegenüber den Strukturformeln Spletts (s. Splett, Ahd.Wb. I,1, XXXf.) erheblich vereinfacht, teils auch modifiziert sind:

- Auf eine Klammerung, mit der Splett die Abfolge der einzelnen Wortbildungsschritte anzeigt, wird verzichtet, da die Strukturanalyse, die der Klammerung zugrunde liegt, automatisch kaum zuverlässig durchzuführen ist. Stattdessen werden Morphemgrenzen durch „|“ angezeigt.
- „+“, das in Spletts Formeln Zusammenrückungsfugen anzeigt, bezeichnet auch Kompositionsfugen.
- Die Wortartkennzeichnung durch A = Adjektiv, S = Substantiv, V = Verb, AD = Adverb unterbleibt bei Kernwörtern grundsätzlich: Statt wS, wS usw. steht stets wX. Bei Suffixen steht das Wortartkennzeichen nur in eindeutigen Fällen (z. B. *-ig* sA), nicht aber z. B. bei *-e* und *-en*.
- Von Präfixen (p) werden abweichend von Splett Verbpartikel (vp) unterschieden.

Die vereinfachte Strukturformel wird durch das Hauptsript aus Lexem2 erzeugt. Zunächst werden die Affixe sukzessive durch ihre Struktursymbole ersetzt. Bei dem danach verbleibenden Lexemsegment muss es sich um das Kernwort oder eine Variante des Kernworts handeln, bei Komposita entsprechend um zwei Kernwörter; z. B. („KW“ steht für Kernwort(variante)):

*ab-grünt-lich*: *ab-* → pl, *-lich* → lsA, verbleibend KW *grünt-* → wX; ergibt die Strukturformel: plwXlsA

*bête-gültig*: *-ig* → lsA, *-e* → ls, verbleibend KW *bët* → wX, KW *gült-* → wX; ergibt die Strukturformel: wXls+wXlsA

*un-ab-ge-zèl(e)t*: *un-* → pl, *ab-* → pl, *ge-* → pl, *-(e)t* → |Vpart2|A; verbleibend KW *zèl*; ergibt die Strukturformel: pl|plwX|Vpart2|A

Tab. 5: Beispiele aus der um Strukturformel und Kernwort erweiterten Lemmaliste.

Lexem, POS	Lexem2	Strukturformel	Kernwort
<i>ab-grünt-lich</i> , adj	<i>ab-grünt-lich</i>	plwXlsA	<b>grünt</b>
<i>bëte-gültig</i> , adj	<i>bët-e-gült-ig</i>	wXls+wXlsA	<b>bët+gült</b>
<i>un-ab-ge-zèl(e)t</i> , adj_part	<i>un-ab-ge-zèl-(e)t</i>	plplplwXIVpart2lA	<b>zèl</b>

### 3.2.6. Konstruktion des umlautbereinigten Kernworts

Sodann erzeugt das Script ein weiteres Feld für das „umlautbereinigte“ Kernwort. Das bedeutet, dass ein ahd. *i*-Umlaut im Kernwort rückgängig gemacht wird, z. B. *ü* → *u*, *üe* → *uo*, *öu* → *ou*. Bei Kernwörtern, die nicht vom ahd. *i*-Umlaut betroffen sind,<sup>3</sup> wird das unveränderte Kernwort als umlautbereinigte Form genommen. Nicht berücksichtigt wird bei diesem Schritt der germ. *i*- und *a*-Umlaut, der erst später einbezogen wird (s. 3.2.7). Enthält das Lexem als Kompositum mehr als ein Kernwort, so wird für jedes Kernwort ein weiterer Datensatz hinzugefügt, der im letzten Feld nur die umlautbereinigte Form dieses Kernworts enthält (s. *bëte-gültig* in Tab. 6).

Tab. 6: Beispiele aus der um das umlautbereinigte Kernwort erweiterten Lemmaliste.

Lexem, POS	Lexem2	Strukturformel	Kernwort	umlautbereinigtes Kernwort
<i>ab-grünt-lich</i> , adj	<i>ab-grünt-lich</i>	plwXlsA	<i>grünt</i>	<b>grunt</b>
<i>bëte-gültig</i> , adj	<i>bët-e-gült-ig</i>	wXls+wXlsA	<i>bët+gült</i>	<b>bët</b>
<i>bëte-gültig</i> , adj	<i>bët-e-gült-ig</i>	wXls+wXlsA	<i>bët+gült</i>	<b>gult</b>
<i>un-ab-ge-zèl(e)t</i> , adj_part	<i>un-ab-ge-zèl-(e)t</i>	plplplwXIVpart2lA	<i>zèl</i>	<b>zal</b>

### 3.2.7. Weitere Alternanzbereinigung: Ablaut, Grammatischer Wechsel, Geminatio

Während bei *grunt* und *zal* mit der Umlautbereinigung auch formal bereits das Kernwort der Wortfamilie extrahiert ist, unterscheidet sich *gult* noch durch Ablaut vom Kernwort *gèlt* seiner Wortfamilie GËLTEN.

Wie lässt sich dieses Problem lösen? Eine globale „Ablautbereinigung“ würde jeden Vokal soweit wie möglich einer Ablautreihe zuordnen und dann

3 Bei mhd. *iu* ist eine Umlautbereinigung nicht sinnvoll, da sonst z. B. auch *liut* (in *liut-e* ‚Leute‘) zu *lüt* umgeformt und *liute* damit der Wortfamilie LÛT zugeschlagen würde.

eine „ablautbereinigte“ Form des Kernworts mit dem grundstufigen Vokal erzeugen, z. B. Kernwort *bôt* → Ablautreihe 2b → ablautbereinigtes Kernwort *bièt*. Dies würde jedoch zu einer Menge von Fehlzuordnungen und untauglichen Kernwörtern führen; so würde *brôt* ‚Brot‘ der Ablautreihe 2b zugeordnet und dazu die abwegige Kernwortform *brièt* erzeugt; entsprechend z. B. *bluot* ‚Blut‘/‚Blüte‘ → Ablautreihe 6 → ablautbereinigtes Kernwort *blat*, wodurch sowohl *bluot* ‚Blüte‘ als auch *bluot* ‚Blut‘ der Wortfamilie BLAT (oder gar \*BLATEN!) von *blat* ‚Blatt‘ zugeordnet würden.

Daher empfiehlt sich ein anderer Weg: Ausgangspunkt ist hierbei, dass die Hauptlieferanten solcher ablautbedingter Kernwortvarianten die Wortfamilien starker Verben sind. Deshalb werden durch ein eigenes Script zunächst möglichst alle erwartbaren Kernwortvarianten erzeugt, die bedingt sind durch

- Ablaut,
- Gramm. Wechsel (z. B. *ziehen* – *zugen*),
- Geminatio u. expressive Verschärfung (z. B. *zücken* < \**tukkjan* < \**tuggjan*)
- germ. *i*-Umlaut (z. B. *siht* < \**seht-i-* zu *sëhen*)<sup>4</sup>
- germ. *a*-Umlaut (z. B. *ge-bot* < \**ga-bud-a-* zu *bièten*)

Unter „Alternanzbereinigung“ sei die Zuordnung aller so entstandenen Kernwortvarianten zum Kernwort der jeweiligen Wortfamilie zusammengefasst, das grundstufig und frei von Gramm. Wechsel, Geminatio und germ. *i*-Umlaut ist. So wird z. B. zu umlautbereinigtem *zuck* (und ebenso auch zu *zôh*, *zug*, *zog*, *ziug*, *zock*) das alternanzbereinigte Kernwort *zieh* 2b erzeugt. Die Variantenliste zu den mhd. starken Verben (und Präterito-Präsentien) enthält aktuell ca. 1740 Varianten. Wünschenswert wäre es, auch die Kernwort-Varianten der Präteritopräsentien in die Liste aufzunehmen. Die Liste kann verbessert werden durch Veränderung des sie erzeugenden Scripts, durch manuelles Löschen oder Hinzufügen von Einträgen ohne Veränderung des Quellcodes.

---

4 Hierher auch *i-e*-Alternanzen, die sich anders erklären, wie z. B. *stich* mit altem schwundstufigem *i* zu „ablautentgleistem“ *stëchen* (Seebold 1970: 467f.); bei *j*-Präsentien wie *sitzen* und *bitten* werden wegen der umgelauteten Grundformen *sitz* und *bitt* als alternanzbereinigte Kernwörter angesetzt.

Tab. 7: Ausschnitt aus der Kernwort-Variantenliste (KW = Kernwort der Wortfamilie).

Variante	KW	Reihe	Variante	KW	Reihe	Variante	KW	Reihe
back	back	6	bar	bër	4	bièg	bièg	2a
bâg	bâg	7	bâr	bër	4	biug	bièg	2a
bann	bann	7	bër	bër	4	bock	bièg	2a
balg	bëlg	3b	bir	bër	4	bog	bièg	2a
bëlg	bëlg	3b	bor	bër	4	boug	bièg	2a
bolg	bëlg	3b	bur	bër	4	buck	bièg	2a
bulg	bëlg	3b	barg	bërg	3b	bug	bièg	2a
ball	bëll	3b	bërg	bërg	3b	bugg	bièg	2a
bëll	bëll	3b	borg	bërg	3b	bièt	bièt	2b
boll	bëll	3b	burg	bërg	3b	bot	bièt	2b
bull	bëll	3b				bôt	bièt	2b
						but	bièt	2b

Diese Variantenliste liest das Hauptsript als externe Hilfsdatei ein und prüft, ob die jeweilige umlautbereinigte Kernwort-Variante in der linken Spalte der Liste vorkommt. Wenn ja, wird ihr das alternanzbereinigte Wortfamilien-Kernwort der rechten Spalte zugeordnet. Mit diesem Kernwort und seiner Ablautreihe werden zwei weitere Felder der Lemmaliste gefüllt. Kommt die umlautbereinigte Kernwort-Variante nicht in der Variantenliste vor, so ist sie zugleich auch die alternanzbereinigte Form und erhält die Ablautreihenanzahl 0, das heißt: sie ist synchron mittelhochdeutsch keiner der Wortfamilien starker Verben zugehörig (vgl. z. B. *tag* von *vièr-tègig* ‚viertägig‘ in Tab. 8).

Tab. 8: Beispiele aus der um das alternanzbereinigte Kernwort erweiterten Lemmaliste.

Lexem, POS	Lexem 2	Strukturformel	KW	umlaut-ber. KW	alternanz-ber. KW	Reihe
<i>bête-gültig</i> , adj	<i>bët-e-gült-ig</i>	wXls+wXlsA	<i>bët+gült</i>	<i>gult</i>	<i>gëlt</i>	3b
<i>be-sëtzære</i> , m	<i>be-sëtz-ære</i>	plwXlsS	<i>sëtz</i>	<i>satz</i>	<i>sitz</i>	5b
<i>viir(e)-ge-zoèhe</i> , n	<i>viir(e)-ge-zoèh-e</i>	plplwXls	<i>zoèh</i>	<i>zôh</i>	<i>zièh</i>	2b
<i>vièr-tègig</i> , adj	<i>vièr-tèg-ig</i>	wX+wXlsA	<i>tèg</i>	<i>tag</i>	<i>tag</i>	0

Im letzten Schritt werden die erzeugten Datensätze der Lemmaliste nach dem alternanzbereinigten Kernwort als erstem und der Ablautreihe als zweitem Sortierschlüssel sortiert. Dadurch werden im günstigen Fall alle Mitglieder einer Wortfamilie zusammengeführt. Als dritter und vierter Sortierschlüssel empfehlen sich das umlautbereinigte Kernwort und das Ausgangslexem.

Tab. 9: Ausschnitt aus der sortierten erweiterten Lemmaliste (Anfang der Wortfamilie DWINGEN ‚zwingen‘).

Lexem, POS	Lexem 2	Strukturformel	KW	umlaut-ber. KW	alternanz-ber. KW	Reihe
<i>be-dwang</i> , m	<i>be-dwang</i>	plwX	<i>dwang</i>	<i>dwang</i>	<i>dwing</i>	3a
<i>be-dwangsals</i> , n	<i>be-dwang-sals</i>	plwXlsS	<i>dwang</i>	<i>dwang</i>	<i>dwing</i>	3a
<i>be-dwènge</i> , n	<i>be-dwèng-e</i>	plwXlS	<i>dwèng</i>	<i>dwang</i>	<i>dwing</i>	3a
<i>be-dwèngen</i> , swv	<i>be-dwèng-en</i>	plwXlV	<i>dwèng</i>	<i>dwang</i>	<i>dwing</i>	3a
<i>be-dwèngēt</i> , part adj	<i>be-dwèng-et</i>	plwXlVpart2lA	<i>dwèng</i>	<i>dwang</i>	<i>dwing</i>	3a
<i>be-dwèng-lich</i> , adj	<i>be-dwèng-lich</i>	plwXlsA	<i>dwèng</i>	<i>dwang</i>	<i>dwing</i>	3a
<i>be-dwèngnisse</i> , f	<i>be-dwèng-nisse</i>	plwXlsS	<i>dwèng</i>	<i>dwang</i>	<i>dwing</i>	3a
<i>dwang</i> , m	<i>dwang</i>	wX	<i>dwang</i>	<i>dwang</i>	<i>dwing</i>	3a

#### 4. Fehlerdiagnose und -korrektur

Bei automatisch erzeugten Ergebnissen ist eine gründliche Fehlerdiagnose in aller Regel unerlässlich – sowohl für eine anschließende Verbesserung des Verfahrens als auch für die notwendige Korrektur seiner Ergebnisse! Bei dem hier vorgestellten Verfahren gibt es zwei Haupttypen von Fehlern:

1) Lexeme sind ihrer Wortfamilie nicht zugeordnet worden

Aufgrund von Sonderformen, die von den beteiligten Scripten nicht erfasst sind oder in der Kernwort-Variantenliste (s. 3.2.7) nicht berücksichtigt sind, oder aus anderen Gründen haben Mitglieder einer Wortfamilie nicht das richtige alternanzbereinigte Kernwort erhalten. Dieser Fehlertyp dürfte in zahlreichen Einzelfällen vorliegen, fällt aber bezogen auf die einzelnen Wortfamilien – zumal die großen – quantitativ meist wohl weniger ins Gewicht. So fehlen bei der Wortfamilie ZIEHEN nur 5 (1,1%) von 452 Fällen, bei SITZEN nur einer von 347 Fällen, bei SNĪDEN nur 5 (3,2%) von 158 Fällen (s. Tab. 10). Die Gründe sind verschieden: In einigen Fällen hat das Script Suffixe im Wortinneren nicht abgetrennt, so *-ig-* in *un-ge-zūhtig-et*, *zūhtig-ære*, *zūhtig-unge* und *-el-* in *snitzel-ære*. Fälle wie *all-zoges*, *-zuges* ließen sich durch eine auf Adverbien beschränkte Abtrennung von *-es* beheben, die daher substantivische Lexeme wie *ackes* ‚Axt‘ oder *nickes* ‚Nix, Krokodil‘ nicht beträfe. In *ge-sēz(ze)* hätten vor der Suffixabtrennung die Klammern getilgt werden müssen. Für *snitt-el* und *vrôn-(e)-snitt-ære* fehlt in der Kernwort-Variantenliste der Eintrag des geminierten *snitt* (*snitt – snid* 1a). Bei *snitterære* und *zwèi-snīdrig* statt der gewöhnlichen Formen *snitære* und *zwèi-snīdig* handelt es sich um vereinzelte Sonderformen.

Vielfach liegt der Grund der Fehlzuordnung aber auch in der Gestalt der Lemmata. So ist nicht selten dasselbe Lexem als Kompositumsglied oder in abgeleiteten Bildungen abweichend von der Form des Simplex angesetzt, z. B.

*iuwel(e)* ‚Eule‘, aber *naht-*, *stein-*, *turn-iule*. Auch an nicht standardkonformen Lemmaansätzen, die gelegentlich unkorrigiert in die Bonner Lemmaliste gelangt sind, kann die Zuordnung scheitern; zwei Beispiele: Lexers Ansatz des präfigierten *ge-efern* (1, 777) stimmt nicht zu dem des Simplex *avern*, *ävern*, *äfern* ‚wiederholen‘ (1, 106); im Findebuch-Ansatz *achselspange* (Gärtner u.a. 1992: 5) verhindert *-chs-* statt *-hs-* die Zuordnung zur Wortfamilie AHSE.

Tab. 10: Falsch zugeordnete Lexeme der Wortfamilien ZIEHEN, SITZEN und SNĪDEN.

Lexem, POS	Lexem 2	Strukturformel	KW	umlaut-ber. KW	alternanz-ber. KW	Reihe
<i>all-zoges</i> , adv	<i>all-zoges</i>	wX+wX	<i>all+zoges</i>	<i>zoges</i>	<i>zoges</i>	0
<i>all-zuges</i> , adv	<i>all-zuges</i>	wX+wX	<i>all+zuges</i>	<i>zuges</i>	<i>zuges</i>	0
<i>un-ge-zühtiget</i> , part adj	<i>un-ge-zühtig-et</i>	plplwXlVpart2lA	<i>zühtig</i>	<i>zuchtig</i>	<i>zuchtig</i>	0
<i>zühtigære</i> , m	<i>zühtig-ære</i>	wXlsS	<i>zühtig</i>	<i>zuchtig</i>	<i>zuchtig</i>	0
<i>zühtigunge</i> , f	<i>zühtig-unge</i>	wXlsS	<i>zühtig</i>	<i>zuchtig</i>	<i>zuchtig</i>	0
<i>ge-sěz(ze)</i> , n	<i>ge-sěz(ze)</i>	plwX	<i>sěz(ze)</i>	<i>sěz(ze)</i>	<i>sěz(ze)</i>	0
<i>zwei-snidrig</i> , adj	<i>zwei-snidr-ig</i>	wX+wXlsA	<i>zwei+snidr</i>	<i>snidr</i>	<i>snidr</i>	0
<i>snittel</i> , n	<i>snitt-el</i>	wXlsS	<i>snitt</i>	<i>snitt</i>	<i>snitt</i>	0
<i>vrôn(e)-snittære</i> , m	<i>vrôn(-e)-snitt-ære</i>	wXls+wXlsS	<i>vrôn+snitt</i>	<i>snitt</i>	<i>snitt</i>	0
<i>snitterære</i> , m	<i>snitter-ære</i>	wXlsS	<i>snitter</i>	<i>snitter</i>	<i>snitter</i>	0
<i>snitzelære</i> , m	<i>snitzel-ære</i>	wXlsS	<i>snitzel</i>	<i>snitzel</i>	<i>snitzel</i>	0

## 2) Lexeme einer Wortfamilie sind einer anderen Wortfamilie zugeordnet worden

Nicht selten erhalten Lexeme zweier Wortfamilien dasselbe alternanzbereinigte Kennwort. Dadurch kommt es zu einer partiellen oder kompletten Fusionierung verschiedener Wortfamilien. Dieses Defizit ist gravierender als die zuvor besprochenen Zuordnungsfehler bei einzelnen Lexemen. Zudem ist es in der Regel weder durch Verbesserungen des automatischen Verfahrens noch durch nachträgliches Ergänzen der Kernwort-Variantenliste zu beheben.

So enthält die Wortfamilie des starken Verbs *WINTEN* 3a Mitglieder mit den Ablautformen *wint*, *want* und *wunt*, denen in der Kernwort-Variantenliste das alternanzbereinigte Kernwort *wint* zugeordnet ist. Daher werden der Wortfamilie *WINTEN* im Rahmen dieses Verfahrens zwangsläufig auch die Mitglieder der Wortfamilien *wint* ‚Wind‘, *wint* ‚Windhund‘ (wendischer Hund) und *wunt* ‚wund‘ einverleibt und müssen bei der späteren manuellen Korrektur als eigene Wortfamilien abgetrennt werden (s. unten).

Ein Extrembeispiel ist die Wortfamilie von md. *ver-mëchen* ‚aufhalten, hindern, schwächen‘, die eigentlich nur aus maximal zwei Lexemen bestehen dürfte: aus dem nur in einigen mitteldeutschen, vor allem mittelfränkischen Quellen belegten *vermëchen* und seiner erschlossenen Basis *mëchen*. Dieser

„Kleinstfamilie“ wird über die als Ablautvariante gewertete Kernwortform *mach*<sup>5</sup> die gesamte Wortfamilie MACHEN zugeschlagen, außerdem über *mich* auch *michelen* ‚vergrößern‘: Von 225 Lexemen sind damit 223 (99%) falsch zugeordnet!

Tab. 11: Beispiele von Fehlzuordnungen zur Wortfamilie MËCHEN.

Lexem, POS	Lexem 2	Strukturformel	KW	umlaut-ber. KW	alternanz-ber. KW	Reihe
<i>ab-got-machære</i> , m	<i>ab-got-mach-ære</i>	plwX+wXlsS	<i>got+mach</i>	<i>mach</i>	<i>mëch</i>	5
<i>all-ge-mëch-liche</i> , adv	<i>all-ge-mëch-liche</i>	wX+plwXlsAD	<i>all+mëch</i>	<i>mach</i>	<i>mëch</i>	5
...	...	...	...	...	...	
<i>mëchen</i> , swv	<i>mëch-en</i>	wXIV	<i>mëch</i>	<i>mëch</i>	<i>mëch</i>	5
<i>ver-mëchen</i> , stv5	<i>ver-mëch-en</i>	plwXIV	<i>mëch</i>	<i>mëch</i>	<i>mëch</i>	5
<i>michel(e)n</i> , swv	<i>mich-el(e)n</i>	wXlsV	<i>mich</i>	<i>mich</i>	<i>mëch</i>	5

Im Rahmen des vorgestellten Verfahrens lässt sich dieses Problem am leichtesten manuell dadurch lösen, dass die Einträge „*mach – mëch* 5“ und „*mich – mëch* 5“ aus der Kernwort-Variantenliste gestrichen werden, s. Tab. 12.

Tab. 12: Korrektur der Fehlzuordnungen zur Wortfamilie MËCHEN.

Lexem, POS	Lexem 2	Strukturformel	KW	umlaut-ber. KW	alternanz-ber. KW	Reihe
<i>ab-got-machære</i> , m	<i>ab-got-mach-ære</i>	plwX+wXlsS	<i>got+mach</i>	<i>mach</i>	<i>mach</i>	0
<i>all-ge-mëch-liche</i> , adv	<i>all-ge-mëch-liche</i>	wX+plwXlsAD	<i>all+mëch</i>	<i>mach</i>	<i>mach</i>	0
...	...	...	...	...	...	
<i>mëchen</i> , swv	<i>mëch-en</i>	wXIV	<i>mëch</i>	<i>mëch</i>	<i>mëch</i>	5
<i>ver-mëchen</i> , stv5	<i>ver-mëch-en</i>	plwXIV	<i>mëch</i>	<i>mëch</i>	<i>mëch</i>	5
<i>michel(e)n</i> , swv	<i>mich-el(e)n</i>	wXlsV	<i>mich</i>	<i>mich</i>	<i>mich</i>	0

Schon die vorstehenden Beispiele haben gezeigt: Die primäre Sortierung nach dem alternanzbereinigten Kernwort als dem Kandidaten für das Wortfamilie-Kernwort ergibt ein Angebot an möglichen Wortfamilien, das einer gründlichen manuellen Korrektur und Nachbearbeitung bedarf. Allgemein besteht diese Nachbearbeitung aus

- der Ergänzung fehlender Wortfamilienangehöriger, die in der sortierten Liste blockartig, aber auch vereinzelt und versprengt vorkommen können;
- der Trennung fälschlich ganz oder partiell fusionierter Wortfamilien.

5 Außerdem stützt sich die Annahme, dass (*ver*)*mëchen* starkes Verb sei, offenbar auf nur einen Beleg des Partizip II *vermëchen* in der hessisch-thüringischen ‚Kaiserchronik‘-Handschrift Heidelberg, Cpg 361: *Der lip ift mir uirmechen. Ich nemac nicht mer gesprechin*, 919.

Die einzelnen Korrekturschritte müssen auf die besonderen Gegebenheiten der fusionierten Wortfamilien abgestellt sein. Dies sei am Beispiel der Wortfamilie WINTEN (mit den inkorporierten Wortfamilien WINT und WUNT) erläutert:

- Der Datensatzblock mit dem alternanzbereinigten Kernwort *wint* wird intern nach dem umlautbereinigten Kernwort sortiert (falls dies nicht schon zuvor der dritte Sortierschlüssel war). Dadurch ergeben sich drei Blöcke mit umlautbereinigtem *want*, *wintund* und *wunt*.
- Der Block mit umlautbereinigtem *want* dürfte komplett zur Wortfamilie WINTEN ‚winden‘ gehören, der *wunt*-Block zu WINTEN und WUNT ‚wund‘ und der *wint*-Block zu WINTEN und WINT ‚Wind‘. Die Lexeme der beiden letzten Blöcke müssen daher manuell aufgeteilt und ihren Wortfamilien zugewiesen werden. Bei den zu WINTEN gehörigen Lexemen ist keine Änderung nötig. Bei den zu WUNT gehörigen Lexemen müssen alternanzbereinigtes Kernwort und Ablautreihe von „*wint* 3a“ zu „*wunt* 0“ geändert werden; bei den zu WINT gehörigen ist lediglich die Reihe in „0“ zu ändern.
- Bei der manuellen Auftrennung kann man sich wortfamilienspezifische Besonderheiten zunutze machen. Zu WINT ‚Wind‘ beispielsweise zählen nahezu alle Lexeme mit *-wint*, m (= stm.) als Kompositionsgrundwort und mit *wint-* als Kompositionserstglied (im Ausgangslexem der Lemmaliste); das ist die Hauptmasse der zu dieser Wortfamilie gehörigen Wörter.
- Ähnlich verhält es sich im *wunt*-Block: Fast alle Lexeme mit (*-wunt*, adj-*wunte*, f oder mit *wunte-*, *wunten-* als Erstglied zählen zur Wortfamilie WUNT.
- Bei einer abschließenden Durchsicht sind alle bislang noch falsch zugeordneten Fällen entsprechend zu korrigieren.
- Anschließend ist neu zu sortieren mit dem alternanzbereinigtem Kernwort als erstem, der Ablautreihe als zweitem und dem umlautbereinigtem Kernwort als drittem Sortierschlüssel.
- Dadurch sind nun jeweils zusammensortiert: 1. die Wortfamilie WINTEN (*wint* 3a) mit 482 Mitgliedern, 2. die Wortfamilie WINT (*wint* 0) mit 75 Mitgliedern und 3. die Wortfamilie WUNT (*wunt* 0) mit 58 Mitgliedern.

Die Beispiele verdeutlichen, dass der benötigte Korrekturaufwand erheblich sein wird. Da sich die vorbereitenden Schritte (s. 3.2) in Grenzen halten und das Weitere bis hin zu der nach den Wortfamilienkernwörtern sortierten Liste weitestgehend automatisch abläuft, beginnt bei dem hier vorgestellten Verfahren mit der manuellen Korrektur die eigentliche Arbeit. Der Hauptaufwand wird dabei nicht in der Kontrolle der sehr großen Wortfamilien wie WINTEN bestehen, deren Zahl überschaubar ist. Ein größeres Problem dürfte

die Masse von 5635 Einzeleinträgen<sup>6</sup> sein, die in der aktuellen sortierten Liste den 4849 potentiellen Wortfamilien mit wenigstens zwei Mitgliedern gegenüberstehen. Zum Vergleich: In Spletts Ahd. Wörterbuch kommen auf ca. 1400 Wortfamilien nur ca. 720 Einzeleinträge (Splett 1993 I,1: LV–LXXVI; I,2: 1209–1240). Bei einer ähnlichen Proportion sollten im Mittelhochdeutschen auf die potentiellen 4849 Wortfamilien nur etwa 2500 Einzelträge kommen. Auch wenn der Anteil der Einzeleinträge, also der familiär isolierten Lexeme, im Mittelhochdeutschen schon wegen des höheren Lehnwortanteils<sup>7</sup> deutlich größer sein dürfte als im Althochdeutschen, muss wohl damit gerechnet werden, dass eine hohe Zahl von isolierten Lexemen aus unterschiedlichen Gründen (s. oben 4.1) ein falsches (das heißt, nicht zielführendes) alternanzbereinigtes Kernwort erhalten hat. Dies zu korrigieren, ist eine mühselige und zeitraubende Aufgabe. Es ist aber zu betonen, dass sich diese Arbeit auch bei jeder anderweitigen umfassenden Wortfamilienuntersuchung des mhd. Wortschatzes nicht umgehen lässt. Bei dem hier vorgestellten Verfahren wird die Korrektur immerhin dadurch erleichtert, dass das Zusammengehörige in der sortierten Liste in der Regel nahe beieinander steht. So wird *èckern* ‚Buchecker, Eichel‘ in *buoch-èckern*<sup>8</sup> über umlautbereinigtes *ackern* in der sortierten Liste zum Nachbarn von *ackeran*<sup>9</sup> ‚Buchecker, Eichel‘ und kann so leicht mit ihm zu einer nur aus diesen beiden Lexemen bestehenden Wortfamilie zusammengeführt werden.

## 5. Ausblick

Nach Abschluss der Korrekturarbeiten lässt sich der Beitrag der mhd. Zeit zur Strukturgeschichte des deutschen Wortschatzes genauer untersuchen. So fragt sich beispielsweise: Wie verändert sich global gesehen der Bestand an Wortfamilien vom Althochdeutschen zum Mittelhochdeutschen und weiter zum Neuhochdeutschen? Welche Wortfamilien sterben wann aus und welche entstehen wann neu? Bleibt dabei der Anteil verbaler und nominaler Kernwörter gleich oder wandelt auch er sich und wenn ja, in welcher Weise?

6 Das sind Lexeme, deren alternanzbereinigtes Kernwort in der Liste nur einmal vorkommt.

7 Nach Stichproben scheint etwa die Hälfte der isolierten Lexeme aus Lehnwörtern zu bestehen.

8 Von Lexer 1, 386 nur in Diefenbachs Glossarium aus einem Vokabular des 15. Jh.s belegt, daher im MWB nicht aufgenommen.

9 In MWB 1, 96 nach einer Sonderform der Urkunde Corpus Nr. 679 (Kloster Andlau/ Elsass 1284) 91,46, 92,9f. kaum zu Recht als *ackeram* statt *ackeran* angesetzt, vgl. Kluge/Seebold (2002: 226 s.v. Ecker). Zum Suffixablaut *-ana-l-ina-* in germ. *\*akr-ana-* vgl. Much (1909: 215f.); daher ist *eckern* (< *\*akr-ina-*) unter Wortbildungsaspekt ein besonderes Lexem und nicht nur Variante von *ackeran*.

Beschränkt man sich zunächst auf einzelne (schon hinreichend korrigierte) Wortfamilien, so lassen sich deren Schrumpfs- und Wachstumsprozesse bereits vorab untersuchen. Als Beispiele dafür mögen abschließend die Wortfamilien ahd. BIOTAN / mhd. BIETEN und ahd. KIOSAN / mhd. KIESEN ‚wählen, prüfen‘ dienen. Vom Althochdeutschen her gesehen setzen sich etwa 53% der Lexeme der Wortfamilie BIOTAN im Mittelhochdeutschen fort; das ist etwas mehr als der Durchschnitt von 47% auch im Mittelhochdeutschen belegten ahd. Lexemen. Die Wortfamilie KIOSAN entspricht mit ca. 48% dagegen dem Durchschnitt. Aus mittelhochdeutscher Warte stellt sich der Vergleich so dar: Nur 17% der mhd. Lexeme haben ahd. Vorläufer, 83% kommen neu hinzu. Davon weicht die Wortfamilie BIETEN mit 13% aus dem Althochdeutschen übernommener und 87% neuer Lexeme nur wenig ab. Bei KIESEN ist dagegen der Anteil schon ahd. belegter Lexeme mit 26% doppelt so hoch wie bei BIETEN und der Anteil neuer Lexeme mit 76% entsprechend deutlich niedriger. Diese Zahlenverhältnisse weisen möglicherweise darauf hin, dass die Wortfamilie BIETEN lebenskräftiger ist als KIESEN: BIETEN hat vom Althochdeutschen zum Mittelhochdeutschen prozentual weniger Lexeme verloren als KIESEN und in mhd. Zeit deutlich mehr Lexeme hinzugewonnen. Bei KIESEN deutet sich in dem unterdurchschnittlichen Lexemzuwachs in mhd. Zeit vielleicht schon ein Vitalitätsverlust an, der auf den fast kompletten Untergang dieser Wortfamilie im Neuhochdeutschen vorausweisen könnte.

## Literatur

- Ganslmayer, Christine. 2012. *Adjektivderivation in der Urkundensprache des 13. Jahrhunderts. Eine historisch-synchrone Untersuchung anhand der ältesten deutschsprachigen Originalurkunden* (Studia Linguistica Germanica 97). Berlin & Boston: de Gruyter.
- Gärtner, Kurt, Christoph Gerhardt, Jürgen Jährling, Ralf Plate, Walter Röll & Erika Timm. 1992. *Findebuch zum mittelhochdeutschen Wortschatz, mit einem rückläufigen Index*. Stuttgart: Hirzel.
- Herbers, Birgit. 2002. *Verbale Präfigierung im Mittelhochdeutschen. Eine semantisch-funktionale Korpusanalyse* (Studien zur Mittelhochdeutschen Grammatik 1). Tübingen: Niemeyer.
- Klein, Thomas, Hans-Joachim Solms & Klaus-Peter Wegera (eds.). 2009. *Mittelhochdeutsche Grammatik. Teil III: Wortbildung*. Tübingen: Niemeyer.
- Kluge, Friedrich. 2002. *Etymologisches Wörterbuch der deutschen Sprache*. Bearb. von Elmar Seebold. 14. Aufl. Berlin & New York: de Gruyter.
- Leipold, Aletta. 2006. *Verbableitung im Mittelhochdeutschen. Eine synchron-funktionale Analyse der Motivationsbeziehungen suffixaler Verbwortbildungen*. Tübingen: Niemeyer.
- Much, Rudolf. 1909. Got. haiþnô. *Zeitschrift für deutsche Wortforschung* 11. 211–218.

- MWB = Gärtner, Kurt, Klaus Grubmüller & Karl Stackmann (eds.). 2006ff. *Mittelhochdeutsches Wörterbuch*. Bd. 1: *a – êvrouwe* (2013). Bd. 2: Doppellieferung 1/2: *êvüegerinne – geværlich* (2014); Doppellieferung 3/4: *gevaterne – hanken* (2017). Stuttgart: Hirzel.
- Ring, Uli. 2008. *Substantivderivation in der Urkundensprache des 13. Jahrhunderts. Eine historisch-synchrone Untersuchung anhand der ältesten deutschsprachigen Originalurkunden*. Berlin & New York: de Gruyter.
- Seebold, Elmar. 1970. *Vergleichendes und etymologisches Wörterbuch der germanischen starken Verben* (Janua linguarum. Series practica 85). Den Haag & Paris: Mouton.
- Splett, Jochen. 1993. *Althochdeutsches Wörterbuch. Analyse der Wortfamilienstrukturen des Althochdeutschen, zugleich Grundlegung einer zukünftigen Strukturgeschichte des deutschen Wortschatzes*. 3 Bde. Berlin & New York: de Gruyter.
- Splett, Jochen. 2009. *Deutsches Wortfamilienwörterbuch. Analyse der Wortfamilienstrukturen der deutschen Gegenwartssprache, zugleich Grundlegung einer zukünftigen Strukturgeschichte des deutschen Wortschatzes*. 18 Bde. Berlin & New York: de Gruyter.

Prof. em. Dr. Thomas Klein  
 Rheinische Friedrich-Wilhelms-Universität Bonn  
 Institut für Germanistik, Vergleichende Literatur- und Kulturwissenschaft  
 Am Hofgarten 22  
 D-53113 Bonn  
[thomas.klein@uni-bonn.de](mailto:thomas.klein@uni-bonn.de)